

硕 士 结 课 作 业

课程名称：法律服务与司法管理技术

论文题目：大数据对法律世界的介入、影响及冲击

姓　　名：陈英熠

学　　院：法学院

专　　业：法律职业伦理

学　　号：2001010114

成　　绩：

大数据对法律世界的介入、影响及冲击

——读《Data-driven law》第八章“Uncovering Big Bias with Big Data: An Introduction to Linear Regression”

引言

周老师带到课堂上与我们相遇的这本书——《Data-driven law：Data analytics and the New Legal Services》（《数据驱动法律：数据分析和新的法律服务》），是对数据在法律世界中与法律的各种互动关系的呈现与研究，作为典型的技术与法律的交叉研究，该书以及该课程均在很大程度上打开了笔者作为纯法学生的学术和专业视野，法律世界不能与世隔绝、亭然独立而存在，势必需要与现实社会的各方因素形成互动与联系，在数字化、大数据等新技术快速发展的当下，数据成为无可避免的、且已在法律之门前徘徊数久的热络门客，数据一步步介入这个沉闷而又保守的法律世界，但法律世界又将如何接纳他？数据将在法律世界中获得何种角色与地位？数据又将在法律世界发挥何种价值？

到底是相对封闭些，坚守自身特有的话语模式，还是适当打开自己，接受其他学科的影响，一直以来都是法学研究时不时面临的选择。比如，经济学之于法学、社会学之于法学、政治学之于法学，等等。本篇浅显的读后报告试图以笔者从本书中获得的对数据的了解出发，尝试对二者关系进行进一步探索。

一、何为法律大数据

什么是大数据？有一本英国学者写的《大数据时代：生活、工作与思维的大变革》，从中大概得知何为大数据。所谓大数据，有三个特征：全样本、混杂性、相关性。其中，最重要的就是全样本。做经验研究的都知道，当样本等于总体时，抽样误差为零。但是，由于财力、人力、分析技术等条件的限制，人们很难拿到全样本。最早，国家为了知道纳税人的实际情况，就发展出各种消减、控制抽样误差的统计技术。而现在，随着计算机技术的发展，人们惊讶地发现，即使面对海量的信息，获取某类现象的全样本也并非完全不可能。基于这种全样本，人们可能更好地了解现实社会中的各种真实。于是，根据这些真实去预测某种现象的发生概率，就更可靠了。可见，大数据并不在于样本绝对量的大小，关键在于“全”

[1]。

对于"什么是法律大数据"，法律界目前也并没有统一明确的说法。套用维克托关于大数据的认识，我们不妨将法律大数据理解为：以一种前所未有的方式，通过对海量法律数据进行分析，对法律问题进行预判，获得巨大价值的产品和服务，或得出新的认知、深刻的观点和主张，法律大数据可能改变法律服务市场及组织机构框架，甚至改变政府与公民的关系。结合课堂上与周老师的探讨与交流，法律大数据并没有一个明确的对数据内容或范畴的边界界定，法律大数据这一概念最终可能成为法律人的一种新的思维方式，通过结合数据分析获得经验并借此认识法律、分析法律的方法论。

二、大数据对法律世界的介入

（一）大数据对法学研究的影响

法律大数据的出现，使得基于法律大数据的司法实践与新型实证研究成为可能，并可能带来法学研究方式的革命性变化。这种可能性源于大数据所具有的独特优势：数据的"全样本性"。大数据通常是特定领域的全面数据，具有数量巨大与内容全面之特性。基于全样本数据的实证研究，能够显著减少传统抽样方法可能导致的误差，增强对研究对象的整体把握，发现传统抽样数据中难以或根本无法获取的信息，带来研究视角、研究素材、研究方法的根本性转变。数据产生、收集、分析的快捷性。"数据分析的速度越来越快，经常在数据刚刚敲进去的时候就可以看到实时的分析结果"，这有助于研究者及时有效地掌握相关法律实践状况的全貌，从而克服传统实证研究方法耗时、滞后的缺陷。数据收集与分析技术的客观性、科学性。

与具有亲历性的传统手工作坊式实证研究"大多是自己收集、整理数据""存在因为研究动机需要而选择性收集、运用数据"不同，海量材料与数据远非"人工作坊时代"研究者所能亲自、逐一地审阅、统计和分析。大数据的收集和分析往往直接依托于数据技术自动处理、完成。在开源条件下，研究过程具有相当的透明度，研究结论可复盘检验，数据收集、分析的客观性、科学性明显增强。特别是，利用不同渠道收集的数据集产生了海量数据，当这些数据聚合到一起，可以对其进行挖掘，并开展更深层次的分析，该深度分析能揭示出各种模式、相关关系，并进行有统计意义的各种预测。这不仅能够开展历时性与变迁性的研究，

[1] 白建军. 大数据对法学研究的些许影响[J]. 中外法学, 2015, 27(1):7.

也能够进行预测性研究与趋势分析，最终促进研究科学水准的提升。

近年来，国内也出现直接利用大量数据展开法学研究的探索，并已经注意到法律大数据所面临的伦理规范等问题。其中，有学者就如何开展大数据法律研究，提出了有启发性的见解。不过，国内的大数据法律研究整体上还处于探索阶段，一些研究缺乏对法律大数据的基本认识，研究方法和过程其实建立在某些误识上。因此，检视大数据法律研究现状，澄清若干误识，对于大数据法律研究的健康开展具有基础性意义。

（二）大数据对法律实践的影响

已经有诸多研究指出，大数据分析在推动国家治理决策和提高国家治理能力方面具有积极意义，特别是在提供公共服务方面。就立法活动本身而言，大数据也可以提供更好的证据支持和正当性证明，特别是在监管立法领域（例如食品药品、生产和环境安全等数据密集型领域）。传统的立法起草过程涉及调查、听证、论证、征集意见、风险评估等方式，其目的都在于尽可能多地获得利益相关方的想法和意见。基于海量数据的分析可以使立法起草者获得更加准确的一手数据，从而避免利益相关人基于自身利益而对立法目标进行的扭曲或忽视，成为科学决策的有力补充。

大数据分析有可能超越了原来基于国家能力不足而导致的事实上的政府与社会和私人领域的边界，从而要求在法律上重新划定边界，并限制某些分析和预测性的使用。从这个意义上讲，技术上的精确性并不能完全取代我们对立法原则和目的的探求，有必要就精准技术带的权力边界扩张和后果进行评估。当然，至少是在目前，由于人口流动和城市地区快速发展变动带来的干扰，在运用信息技术加强社会治安防控体系建设的大思路下，一些城市的警方还只能根据大数据进行粗略的预测。例如，北京市怀柔区警方通过运用大数据、云计算和科学分析模型，整合历年案件信息建立了犯罪数据分析和趋势预测系统，能够自动预测犯罪趋势，指导警力投入。除了入室盗窃、诈骗、抢劫等八类犯罪外，警方还将系统信息录入范围拓展到治安案件、交通、火灾事故，以及其他公共管理事件，远未到针对个体的精细治理。[2]但可以预见，随着数据库的扩大和算法的改进，大数据将在未来的智慧城市治理和风险预防中发挥越来越大的作用。

目前，中国利用大数据开展的法律实践方兴未艾。例如：基于司法公开而大力推进的裁判文书上网工作；依托大数据技术建立犯罪信息判断和趋势预测；运

[2] 左卫民.迈向大数据法律研究[J].法学研究,2018,40(04):139-150.

用大数据建设“检察大数据标准体系、应用体系、管理体系、科技支撑体系”；利用大数据建立案件权重系数和评价指标体系，确定法官工作量，并进行科学的员额分配、案件分流；基于大数据开展的多种法律人工智能实践，尝试如类案推荐、量刑辅助与偏离预警等应用。其中，裁判文书大规模上网，使得中国第一次有了全国性、公开的、细节化的法律数据。但总体而言，目前国内对于法律大数据的实践性运用还相对有限，具体运用并不普遍，在一定程度上呈现出“话语热、实践冷”的现象：一方面，应用主体范围有限，主要集中在少数司法机关、法律数据公司；另一方面，应用领域相对较窄、实际运用较少，主要集中在类案检索、法律文书草拟、文书智能纠错等辅助办案方面。

三、大数据对法律世界的冲击与挑战

（一）数据安全与信息保护问题——对个人隐私的挑战

面对大数据时代对包含个人信息的数据的大量的、普遍的采集、存储、处理和流转，个人信息保护中蕴含的人格利益与信息自由流通和应用中蕴含的巨大经济利益、公共安全和公共健康等公共利益之间形成了巨大的张力[3]。这种矛盾和利益冲突在大数据时代将被进一步激化。大数据产生的利益应该如何分配？大数据带来的个人信息被滥用的隐患应该如何规制和防控？面对利益冲突应该如何设置利益平衡点?

（二）算法歧视与“黑箱”问题——对公平正义的威胁

学者们和政府愈来愈意识到算法歧视的广泛存在，并认为这些新型歧视能引发一系列的社会问题、伦理问题和法律问题。算法自动化决策为人们带来方便的同时，也可能因其决策过程的不透明和信息不对称而对某些群体造成歧视。实践中，算法歧视主要表现为偏见代理的算法歧视、特征选择的算法歧视和大数据杀熟三种基本形态。Solon Barocas 和 Andrew D.Selbst 就认为，算法自动化决策可能给某些群体带来歧视性的不利结果。2014 年，白宫发布了一份题为《大数据：抓住机遇，保护价值》（Podesta 等，2014）的报告。该报告认为，受数据来源的特定性和算法设计者主观意图的影响，算法自动化决策往往会对申请者的工作、教育或信贷构成隐性偏见。这种结果可以自我强化，从而系统性地减少个人获得信贷、就业和教育的机会，这会使他们的处境恶化并在未来的申请中处于不

[3] 张吉豫.大数据时代中国司法面临的主要挑战与机遇——兼论大数据时代司法对法学研究及人才培养的需求[J].法制与社会发展,2016,22(06):52-61.

利地位。

四、结论

可以看到大数据开始在法律服务和法学研究上展现其影响力。本文并不提倡一种数据迷信，大数据带来的不会是完全革命性的改变，而只是对决策提供更加精细化的智力支持，使决策和法律执行更有效率和针对性。

一个更加开放的问题是，法律的未来是否会被无尽的数据和算法取代？十多年前就有人提出网络空间中“代码就是法律”的断言。在大数据时代，数据及其形塑数据价值的算法变成了新的网络架构，在貌似客观地分析海量数据的同时，也将这种描述性的遍历性事实作为一种规范性规则施加给每一个人，隐蔽在背后的仍然是不同组织和力量试图挖掘数据价值的计划，尤其是当大数据被广泛应用在私人交易的场合。更进一步，这种网络架构完全可以比现有法律更加精细，从而超越代议制立法机关提出新的规则，而不受到任何制约；当算法复杂到一定程度而无法被人直观理解，形成“黑箱”的时候，这将是危险的。因此，大数据分析的预测性使用不仅不应当取代现有法律，而且还要受到民主机制的制约和监督，即受到民主立法的控制。法律不同于技术架构的要点在于它既是社会中主流价值观的反映，也是多元价值和利益妥协的产物，但数据挖掘只服务于政治或商业的单一利益。在法律实践和研究应用大数据的过程中，我们应当注意这一差别，从而尽可能让大数据服务于公共利益，更好地贴近法律实践。

参考文献：

[1] 左卫民.迈向大数据法律研究[J].法学研究,2018,40(04):139-150.

[2] 郑戈.算法的法律与法律的算法[J].中国法律评论,2018(02):66-85.

[3] 梅夏英.数据的法律属性及其民法定位[J].中国社会科学,2016(09):164-183+209.

[4] 梁吉业,冯晨娇,宋鹏.大数据相关分析综述[J].计算机学报,2016,39(01):1-18.

[5] 张吉豫.大数据时代中国司法面临的主要挑战与机遇——兼论大数据时代司法对法学研究及人才培养的需求[J].法制与社会发展,2016,22(06):52-61.